

Berkeley Lab and Computing Sciences at Berkeley Lab

Horst D. Simon



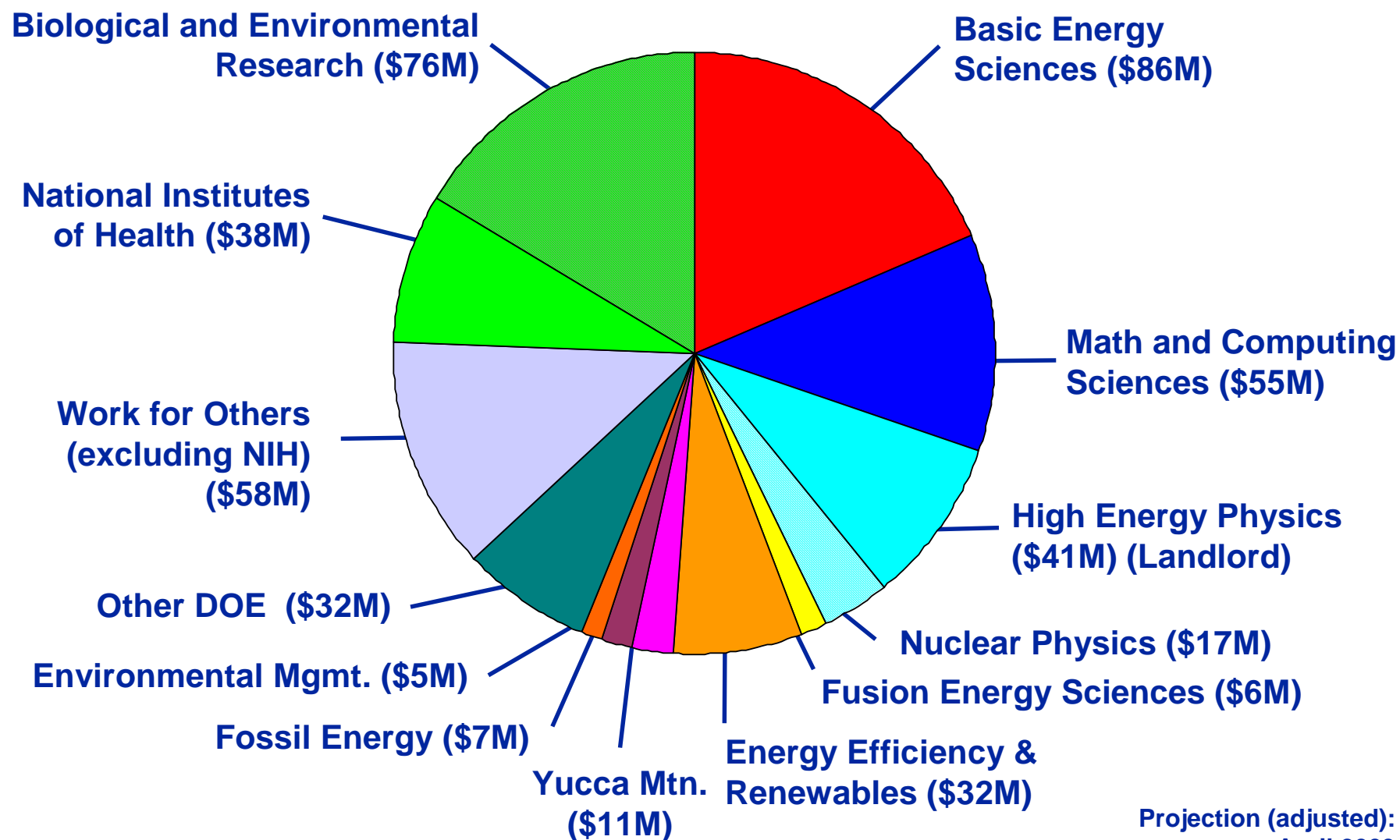
**Coalition of Academic Scientific Computation (CASC)
August 6, 2003**

**Founded in 1931 on the Berkeley Campus
Moved to Current Site in 1940**



LAWRENCE BERKELEY NATIONAL LABORATORY

FY 2003 Budget: \$464M



Projection (adjusted):
April 2003

An aerial photograph of the Lawrence Berkeley National Laboratory campus, showing various research buildings, green spaces, and surrounding hills. The text is overlaid in yellow.

Nanoscience

**Scientific Discovery
through
Advanced
Computing**

**Observing Matter
and Energy in the
Universe**

**Berkeley Lab's
Scientific Vision**

**New Energy Systems
and Environmental
Solutions**

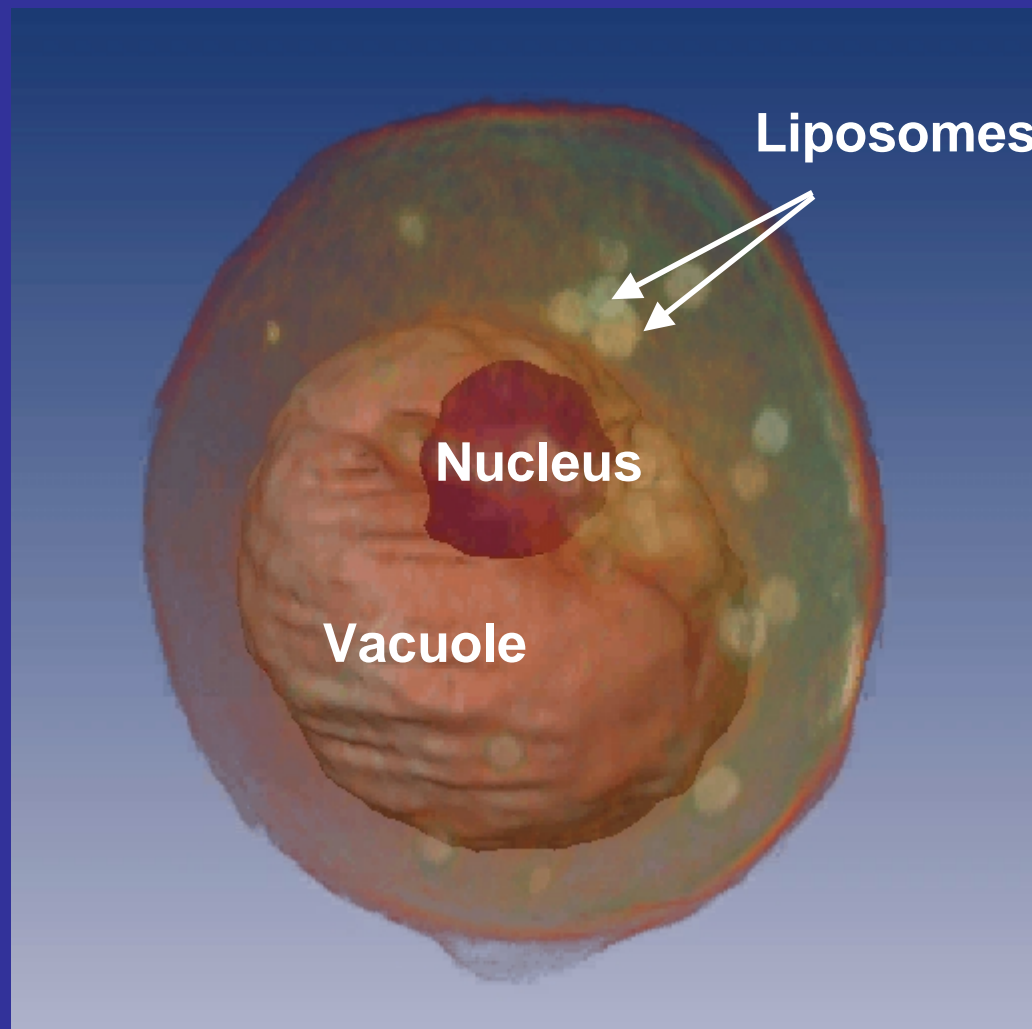
**Quantitative
Biology**

Tenth Anniversary of the Advanced Light Source



LAWRENCE BERKELEY NATIONAL LABORATORY

Soft X-ray Microscopy at the ALS: Tomographic Reconstruction of Single Cell



- Natural contrast x-ray tomogram
- Resolution on order of 50 nanometers
- Reveals yeast subcellular structures

Carolyn Larabell
and Mark Le Gros

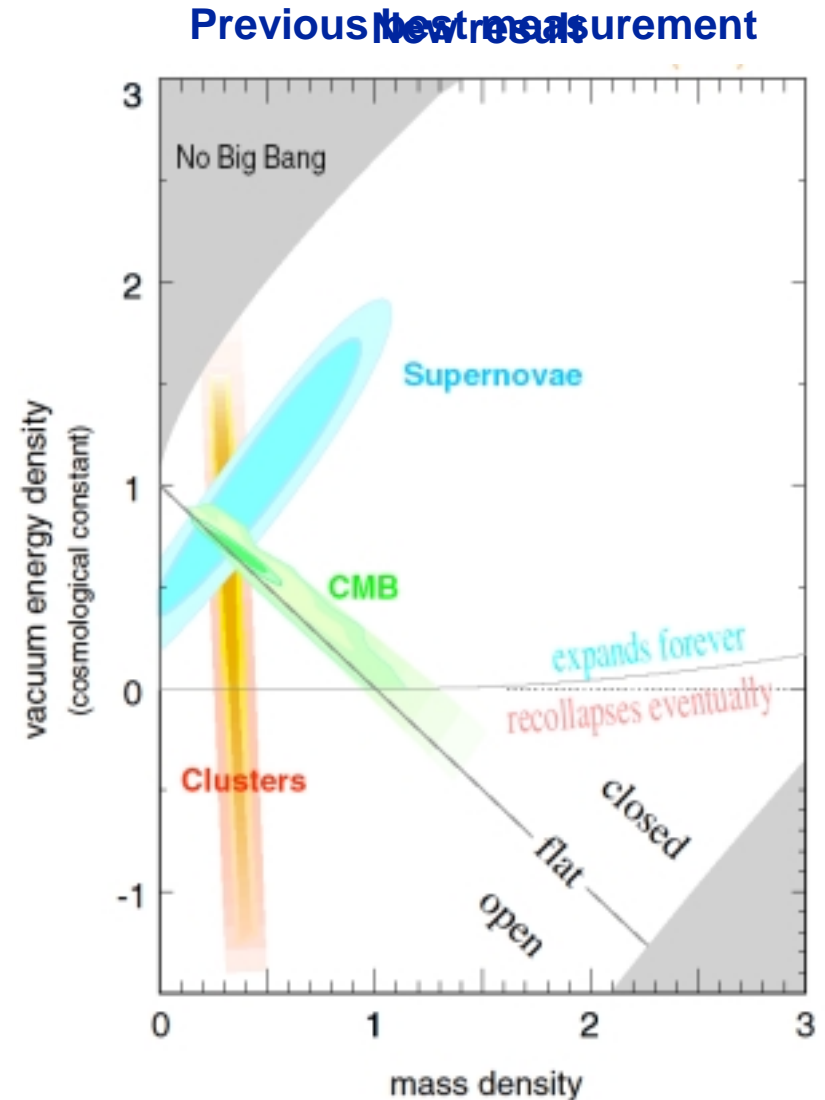
1 μm

New Results Further Confirm Accelerating Universe



- Largest set (11) of distant Type Ia Supernovae detailed with Hubble Telescope
- More precise measurement of cosmological parameters
- Improved spectra demonstrates that host-galaxy dust is not a major systematic bias
- Set indicates improvement anticipated for supernovae to be observed by SNAP

Saul Perlmutter, Greg Aldering,
Rob Knop, et al.



SuperNova/Acceleration Probe Mission

2003

2005

2007

2009

2011

2013

2015

**\$8.4M in
President's
'04 Budget**

R&D

**Design/fabrication
and assembly**

Launch

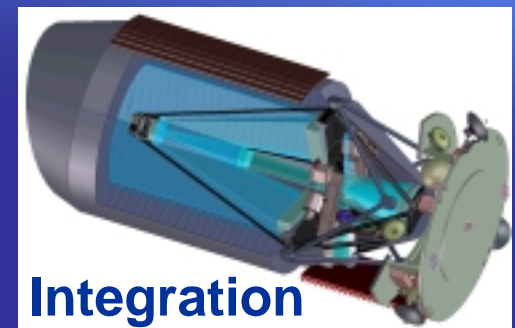
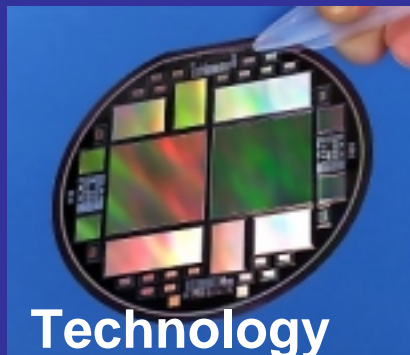
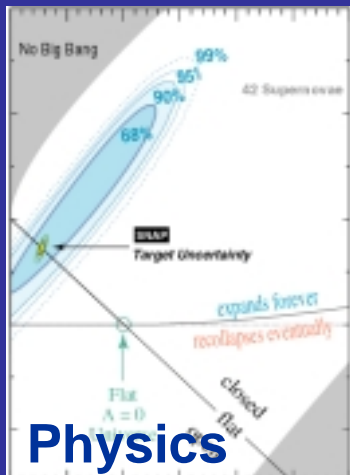
**Observation
and data
analysis**

Discoveries

CD-0

CD-1

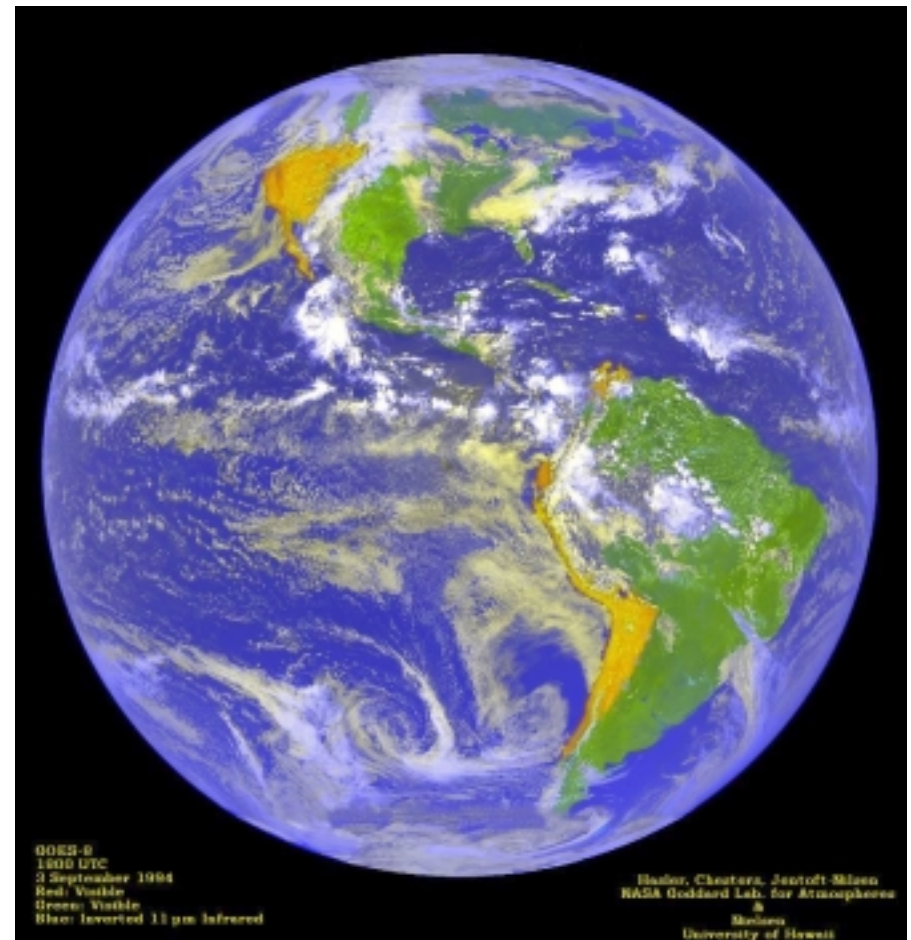
CD-2



Sequencing the Biosphere



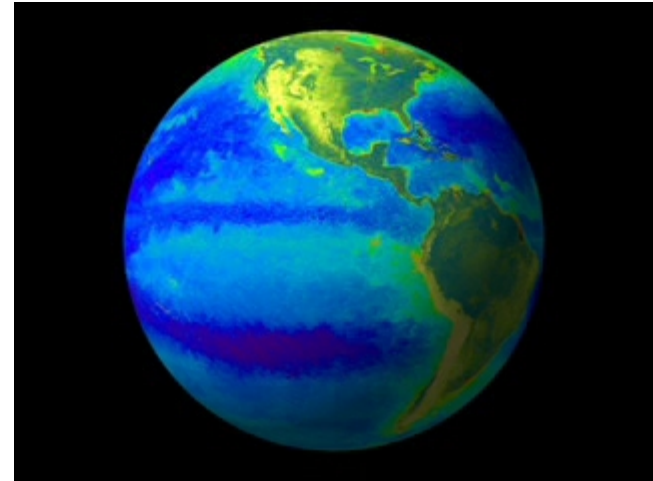
- Less than 1% of microbes are culturable
- Many unculturables live in diverse interdependent consortia
- Aim:
 - Recover genome-scale sequences and reveal metabolic capabilities
 - Understand action of microbes at molecular level
 - Determine structure of microbial populations



Genomics in Support of DOE and National Scientific Missions



- Genomic diagnosis of environment
- JGI as a National User Facility
 - 40% for DOE missions
 - 60% for scientific community
- Department of Agriculture: crop pathogens
- Environmental Protection Agency: environmental indicator species



Sudden Oak
Death



Fathead
Minnow

Joint Genome Institute to Become National User Facility

2003

**National user
facility
operation**

2005

**Increase
sequencing
throughput: catalog
full communities**

2007

2009

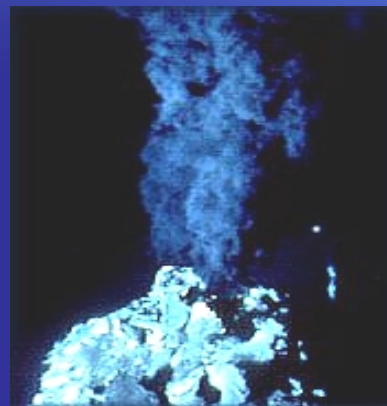
**Expand collection
of biosphere
genomic
repertoire**

2011

2013

**Sequence and
assemble planet's
microbial
diversity**

2015



Molecular Foundry Status



- Design approved by Regents, April 2003
- Successful independent external review completed, June 2003
- Budget:
 - \$6.8M for FY 2003
 - Total project cost: \$85M



Molecular Foundry

2003

2004

2005

2006

2007

2008

2009

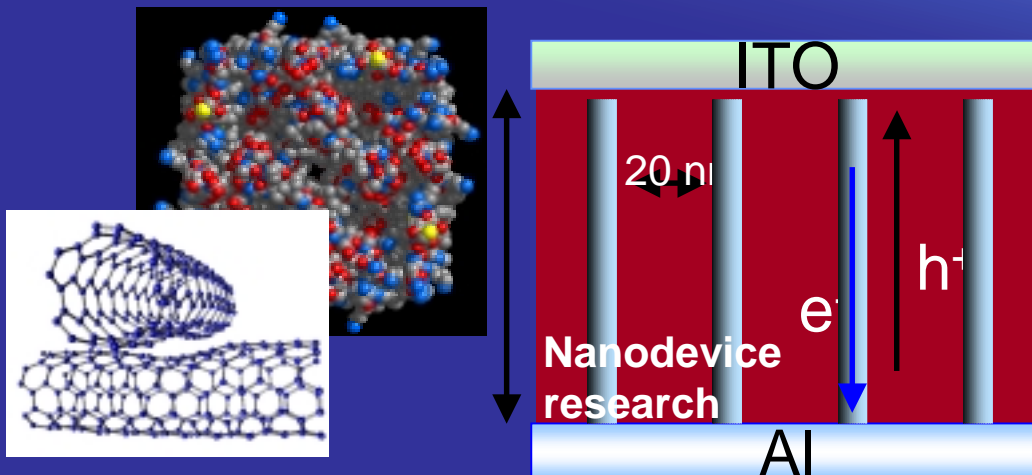
Engineering design
and start construction

Conventional
facilities
complete

Molecular
Foundry
begins
operation

Nanomaterials
and component
design and
fabrication

Hire scientific staff and ramp up user program



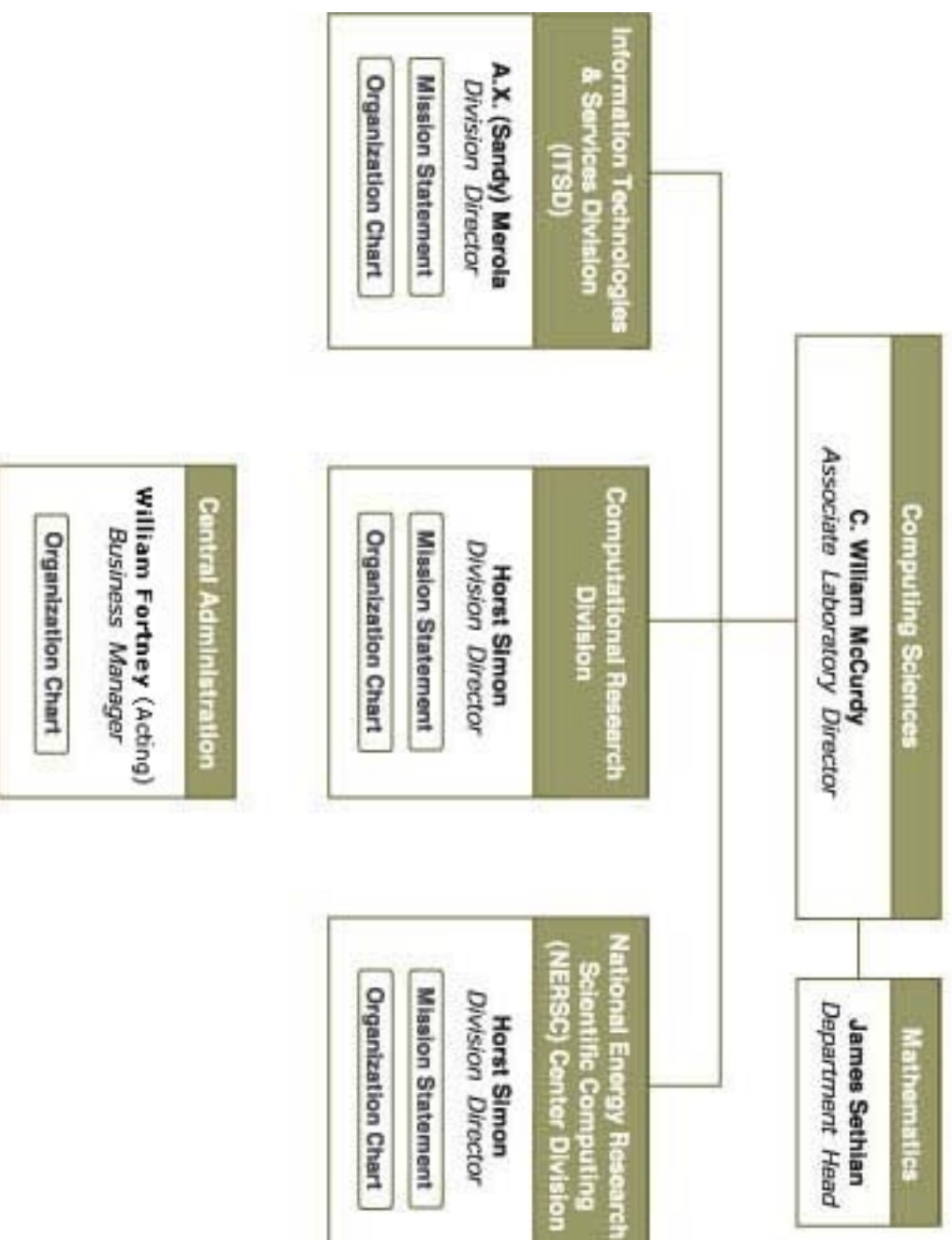
**Applied
Mathematics**

**National Energy
Research Scientific
Computing Center**

**Scientific Discovery
through
Advanced Computing**

**Computational
Science
Research**

**Science Grid and
Energy Sciences
Network**



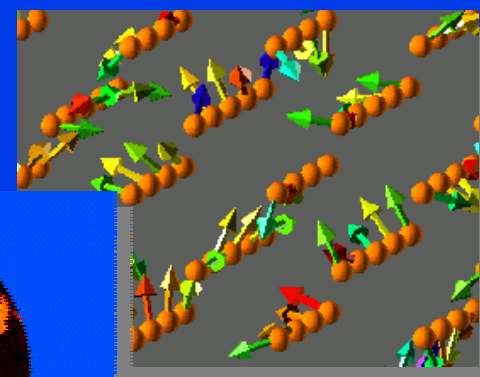
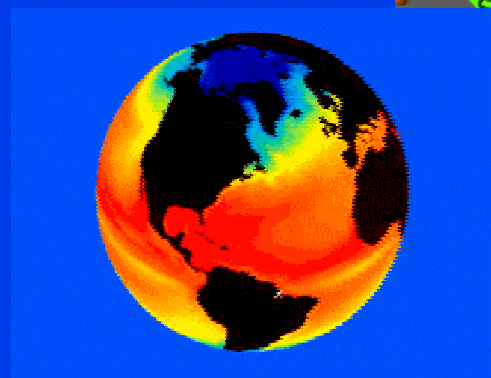
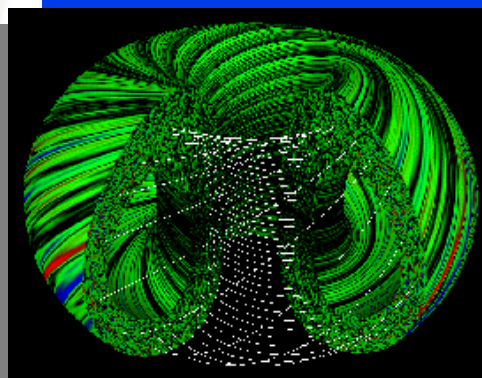
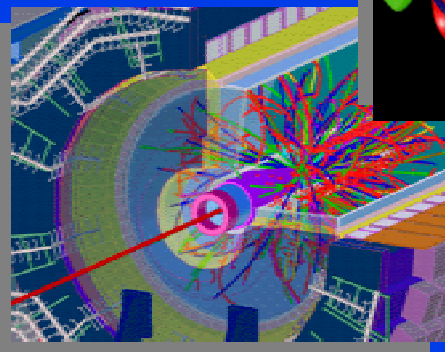
NERSC Center Division

Horst D. Simon, Division Director

**William T.C. Kramer, Division Deputy and
Facility General Manager**

National Energy Research Scientific Computing Center

- Serves all disciplines of the DOE Office of Science
- ~2000 Users in ~400 projects
- Focus on large-scale computing



NERSC Center Overview



- **Funded by DOE, annual budget \$28M, about 65 staff**
- **Supports open, unclassified, basic research**
- **Located in the hills next to University of California, Berkeley campus**
- **Close collaborations between university and NERSC in computer science and computational science**
- **Close collaboration with about 125 scientists in the Computational Research Division at LBNL**



Upgraded NERSC 3E Characteristics



- The upgraded NERSC 3E system has
 - 416 16-way Power 3+ nodes with each CPU at 1.5 Gflop/s
 - 380 for computation
 - 6,656 CPUs – 6,080 for computation
 - Total Peak Performance of 10 Teraflop/s
 - Total Aggregate Memory is 7.8 TB
 - Total GPFS disk will be 44 TB
 - Local system disk is an additional 15 TB
 - Combined SSP-2 is greater than 1.238 Tflop/s
 - NERSC 3E is in full production as of March 1, 2003
 - nodes arrived in the first two weeks of November
 - Acceptance end of December 2002
 - 30-day availability test near completed Feb. 2003
 - In full production March 1, 2003



TOP500 List of Most Powerful Computers



Rank	Manufacturer	Computer	R_{\max} [TF/s]	Installation Site	Country	Year	Area of Installation	# Proc
1	NEC	Earth-Simulator	35.86	Earth Simulator Center	Japan	2002	Research	5120
2	HP	ASCI Q, AlphaServer SC	13.88	Los Alamos National Laboratory	USA	2002	Research	8192
3	Linux Network/ Quadrics	MCR Cluster	7.63	Lawrence Livermore National Laboratory	USA	2002	Research	2304
4	IBM	ASCI White SP Power3	7.3	Lawrence Livermore National Laboratory	USA	2000	Research	8192
5	IBM	Seaborg SP Power 3	7.3	NERSC Lawrence Berkeley Nat. Lab.	USA	2002	Research	6656
6	IBM/Quadrics	xSeries Cluster Xeon 2.4 GHz	6.59	Lawrence Livermore National Laboratory	USA	2003	Research	1920
7	Fujitsu	PRIMEPOWER HPC2500	5.41	National Aerospace Laboratory of Japan	Japan	2002	Research	2304
8	HP	rx2600 Itanium2 Cluster Qadrics	4.88	Pacific Northwest National Laboratory	USA	2003	Research	1536
9	HP	AlphaServer SC ES45 1 GHz	4.46	Pittsburgh Supercomputing Center	USA	2001	Academic	3016
10	HP	AlphaServer SC ES45 1 GHz	3.98	Commissariat a l'Energie Atomique (CEA)	France	2001	Research	2560

Science of Scale: Cosmic Microwave Background Data Analysis



- PI: Julian Borrill, LBNL & UC Berkeley
- Code: Maximum likelihood angular power spectrum estimation (MADCAP)
- Kernel: ScaLAPACK
- Performance: 750 Mflop/s per processor (50% of peak)
- Scalability:
- 0.78 Tflop/s on 1024 proc
- 1.57 Tflop/s on 2048 proc
- 3.02 Tflop/s on 4096 proc

Cosmic Microwave Background Data Analysis (cont.)



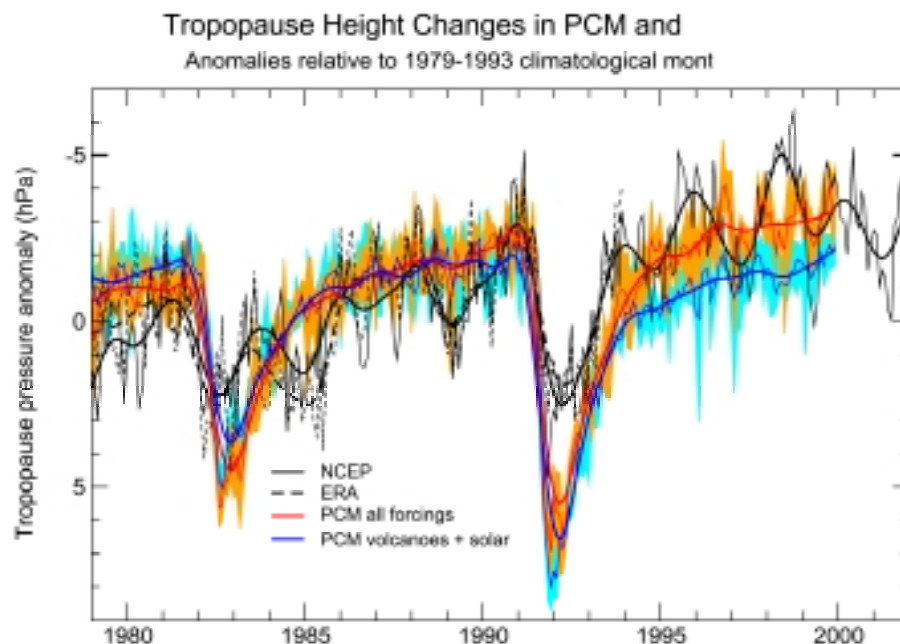
Recent accomplishments:

- **MADCAP** extended to enable simultaneous analysis of multiple datasets and CMB polarization – the new frontier.
- **MADCAP** was rewritten to exploit extremely large parallel systems, allowing near-perfect scaling from 256 to 4,096 processors.
- **MADCAP++** is being developed using approximate methods to handle extremely large datasets for which matrix multiplications are impractical, such as will be generated by the **PLANCK** satellite.
- Recent results from NASA's **WMAP** satellite observations of the whole CMB sky confirm **MADCAP** analyses of previous partial-sky balloon datasets.

Investigating Atmospheric Structure Changes with PCM



- The tropopause is that height demarking the troposphere and the stratosphere.
 - Below the tropopause, the temperature cools with altitude.
 - Above the tropopause, the temperature warms with altitude.
- A diagnostic that is robust to El Nino but sensitive to volcanoes.
- An indicator of the total atmospheric heat content
- Changes in natural forcings alone (blue) fail to simulate this feature of the atmosphere, but natural + anthropogenic changes (orange) do



Santer et al. Figure 1

New Results in Climate Modeling



- Recent improvements in hardware have reduced turnaround time for the Parallel Climate Model
- This has enabled an unprecedented ensemble of numerical experiments.
 - Isolate different sources of atmospheric forcing
 - Natural (solar variability & volcanic aerosols)
 - Human (greenhouse gases, sulfate aerosols, ozone)
- Data from these integrations are freely available to the research community.
 - By far the largest and most complete climate model dataset
 - www.nersc.gov/~mwehner/gcm_data

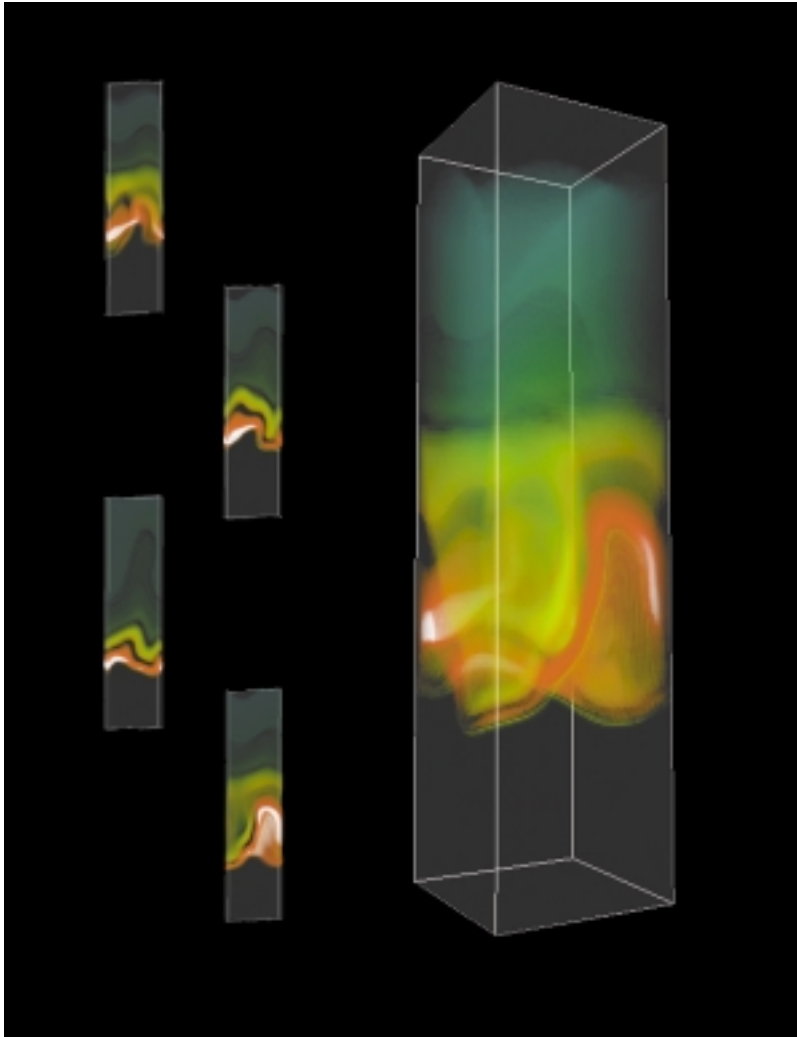
Computational Research Division

Horst D. Simon, Division Director

Juan C. Meza, HPC Research Dept. Head

**William Johnston, Distributed Systems
Dept. Head**

Computational Research Division

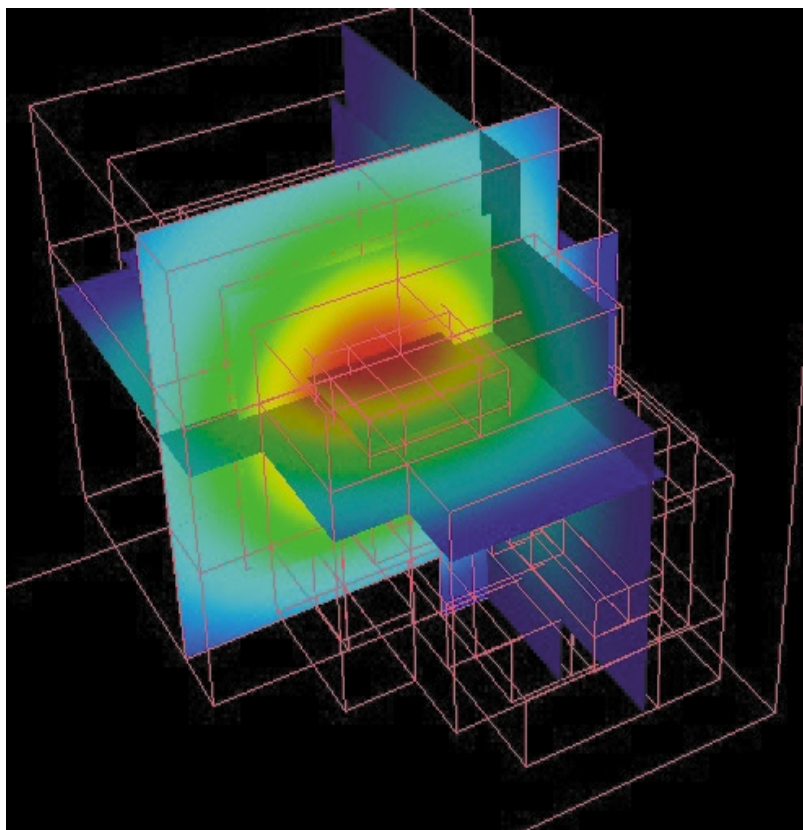


Horst Simon, Division Director

The Computational Research Division(CRD) creates computational tools and techniques that enable scientific breakthroughs, by conducting applied research and development in computer science, computational science, and applied mathematics. CRD consists of two departments:

- 1. High Performance Computing Research**
 - 2. Distributed Systems**
- About \$5M of SciDAC funded projects**

High Performance Computing Research Department (HPCRD)



Juan Meza, Department Head

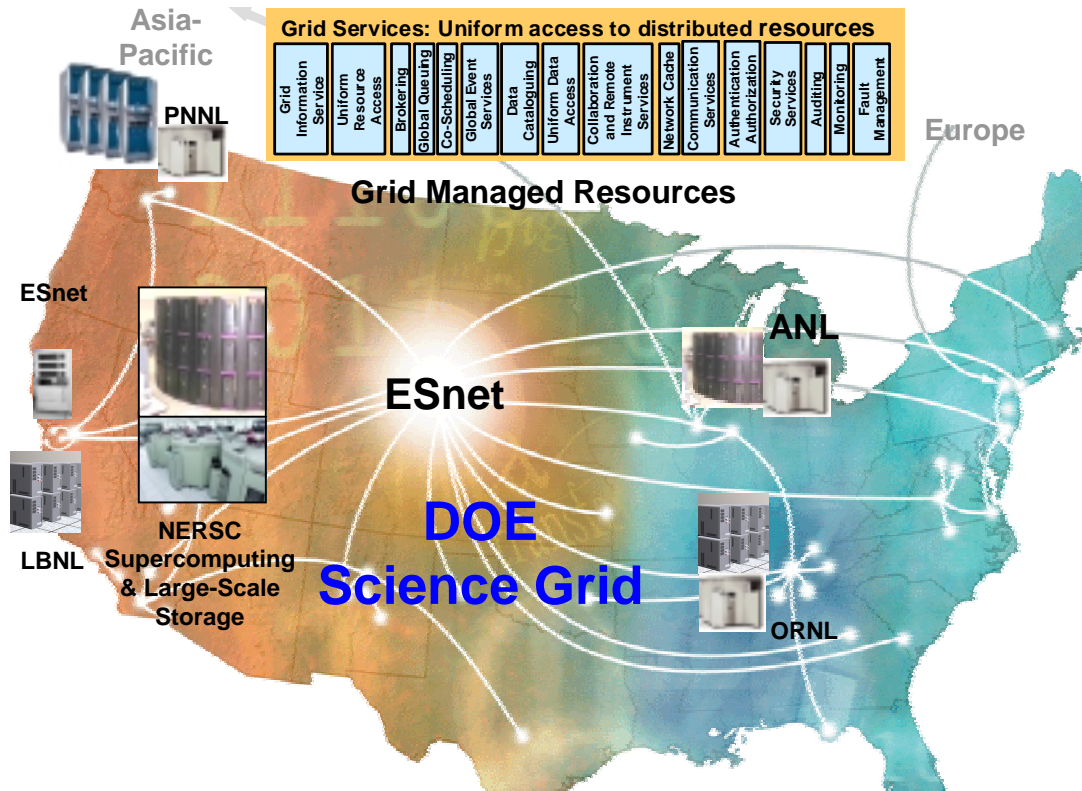
Groups:

- Applied Numerical Algorithms
- Center for Computational Sciences and Engineering
- Future Technologies
- Imaging and Informatics
- Scientific Computing
- Scientific Data Management
- Visualization

Total Staff: 108

The High Performance Computing Research Department conducts research and development in mathematical modeling, algorithmic design, software implementation, and system architectures, and evaluates new and promising technologies.

Distributed Systems Department



William Johnston, Department Head

Deb Agarwal, Department Deputy

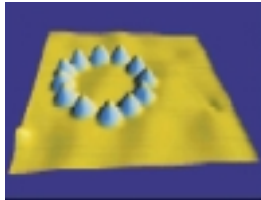
Groups:

- Collaboration Technologies
- Data Intensive Distributed Computing
- Network Technologies
- Secure Grid Technologies

Total Staff: 25

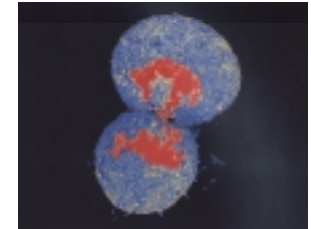
The Distributed Systems Department researches and develops software components that allow scientists to address complex and large-scale computing and data analysis problems in a distributed environment such as the DOE Science Grid.

Computational Science Mission



**nano
systems**

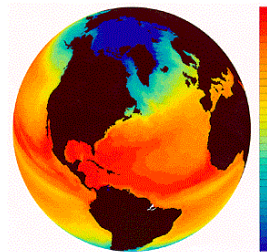
**biological
systems**



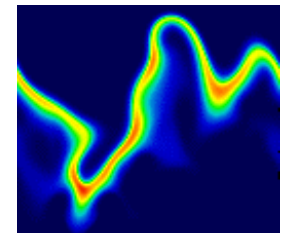
**The Computational Research
Division is engaged in
computational science
collaborations, creating tools and
techniques that will enable
computational modeling and
simulation, and lead to new
understanding in areas such as**



**astrophysics
simulation**



global climate



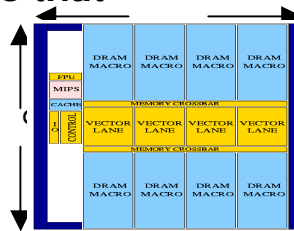
**combustion
processes**

Computer Science and Applied Mathematics Mission

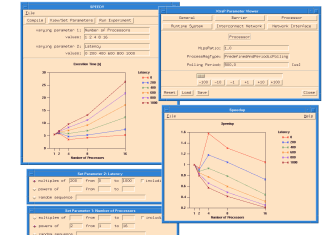


The Computational Research Division is engaged in basic and applied research addressing the following questions of fundamental importance to enabling progress in our ability to use computing and networking technology

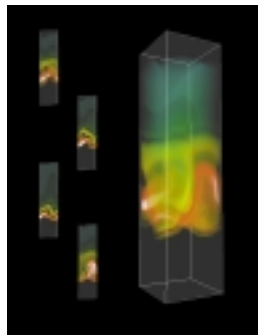
- develop computer architectures that most suitable for scientific applications and measure their effectiveness for science



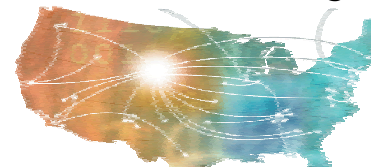
- research in algorithms and development of software tools for these new architectures



- research in algorithms, and development of software tools, and technology in data management, analysis, and visualization



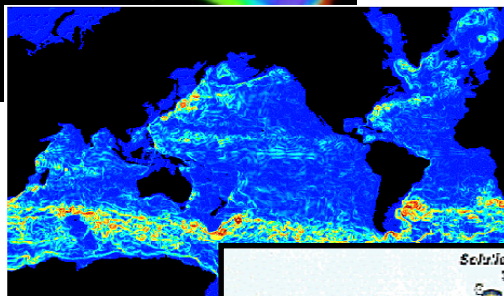
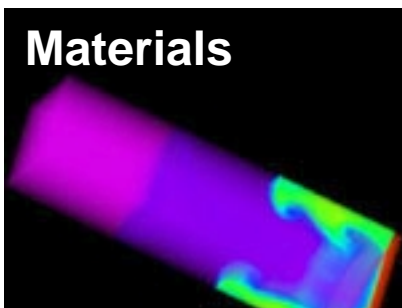
- research in networking and distributed computing, and development of grid middleware and collaboration technologies



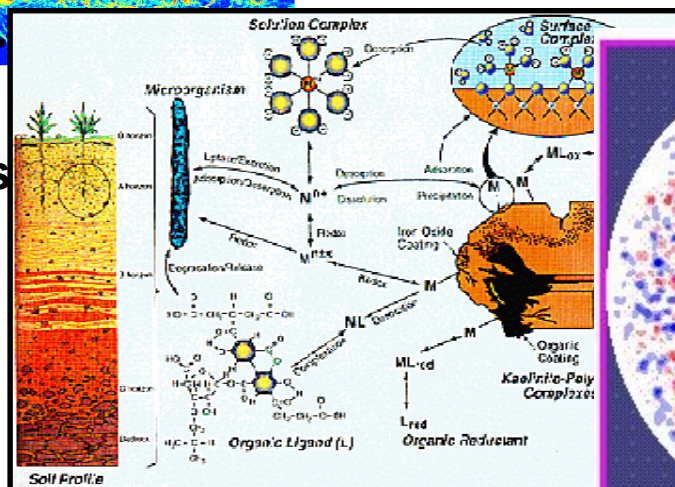
Scientific Discovery Through Advanced Computing – Projects in the Computational Research Division

Scientific Discovery Through Advanced Computing

Materials

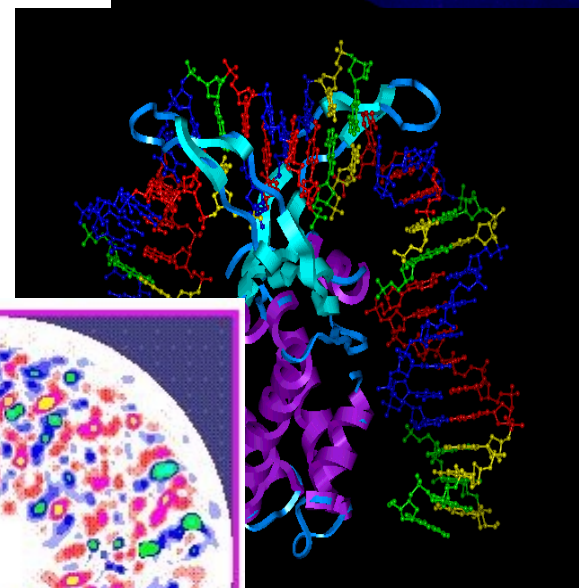
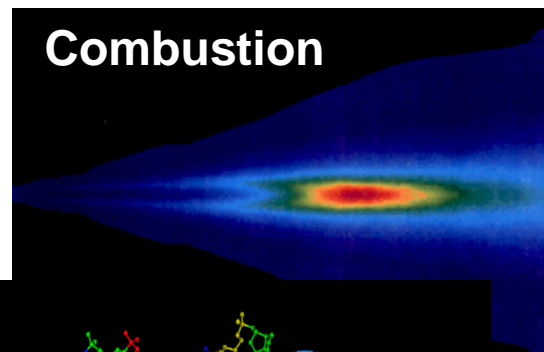


Global Systems

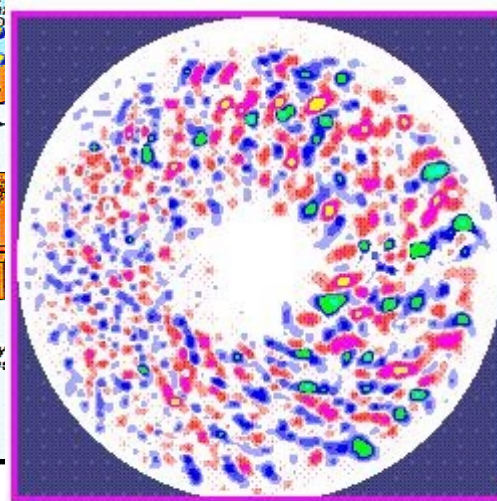


Subsurface Transport

Combustion



Health Effects, Bioremediation

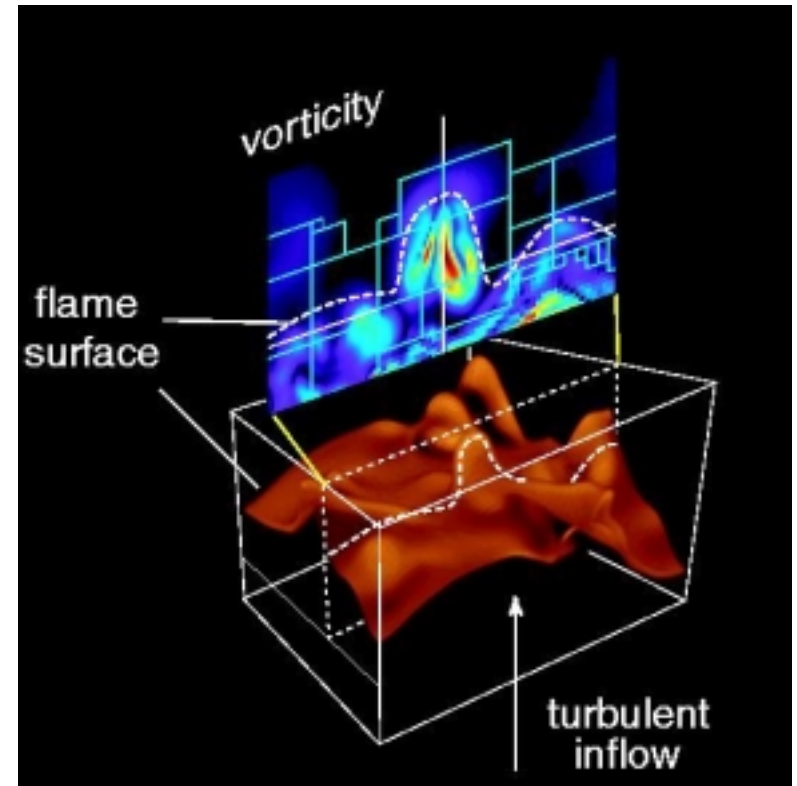


Fusion Energy

Applied Partial Differential Equations ISIC

Developing a new algorithmic and software framework for solving partial differential equations in core mission areas, such as accelerator physics, magnetic fusion, and combustion.

- New algorithmic capabilities with high-performance implementations on high-end computers:
 - Adaptive mesh refinement
 - Cartesian grid embedded boundary methods for complex geometries
 - Fast adaptive particle methods
- Close collaboration with applications scientists
- Common mathematical and software framework for multiple applications



Participants: LBNL (J. Bell, P. Colella), LLNL , Courant Institute, Univ. of Washington, Univ. of North Carolina, UC Davis, Univ. of Wisconsin.

2003 SIAM/ACM Prize in Computational Science and Engineering

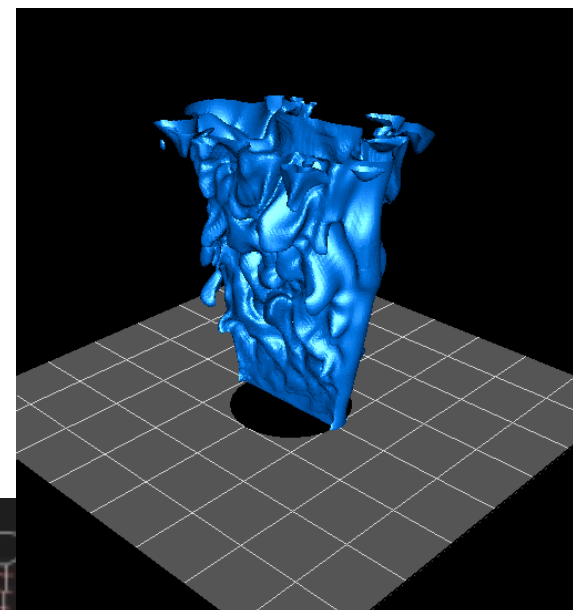


Awarded by the Society for Industrial and Applied Mathematics (SIAM) and the Association for Computing Machinery (ACM)

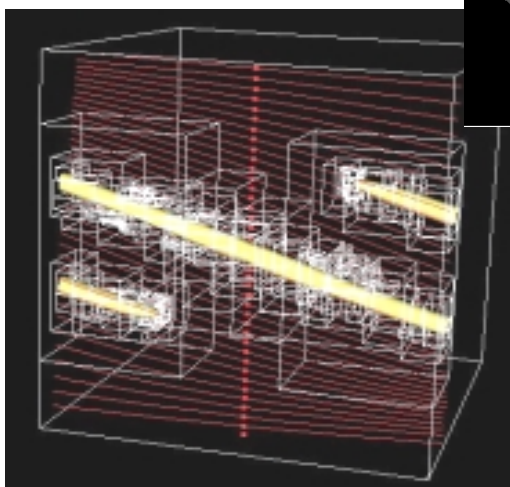


**John
Bell**

**Chemistry and
turbulence in
methane
combustion**



**Phillip
Colella**



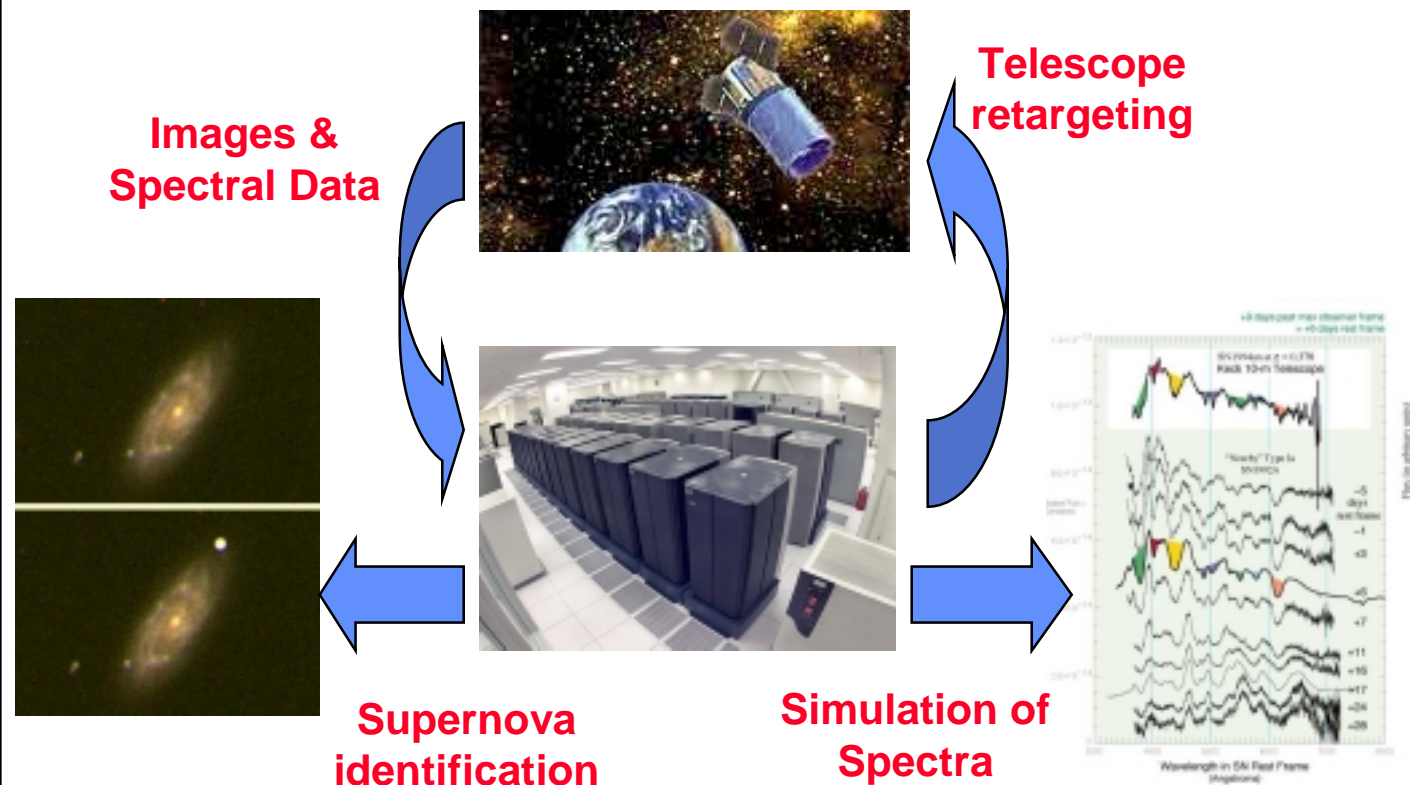
**Three-dimensional adaptive
mesh calculation of pellet
injection into a plasma**

The DOE Science Grid: A New Type of Infrastructure

Early Example:
Access Grid –
Multi-site, high-end
collaboration



Lab, University, and Industrial partners
collaborate using unique instruments, large
amounts of data, and computing facilities at
multiple sites.



Example: The Supernova Factory (SNAP)

Participants: LBNL, ANL, ORNL, PNNL (and NERSC)

Projects in Computational Research Division

--

Bringing all resources together



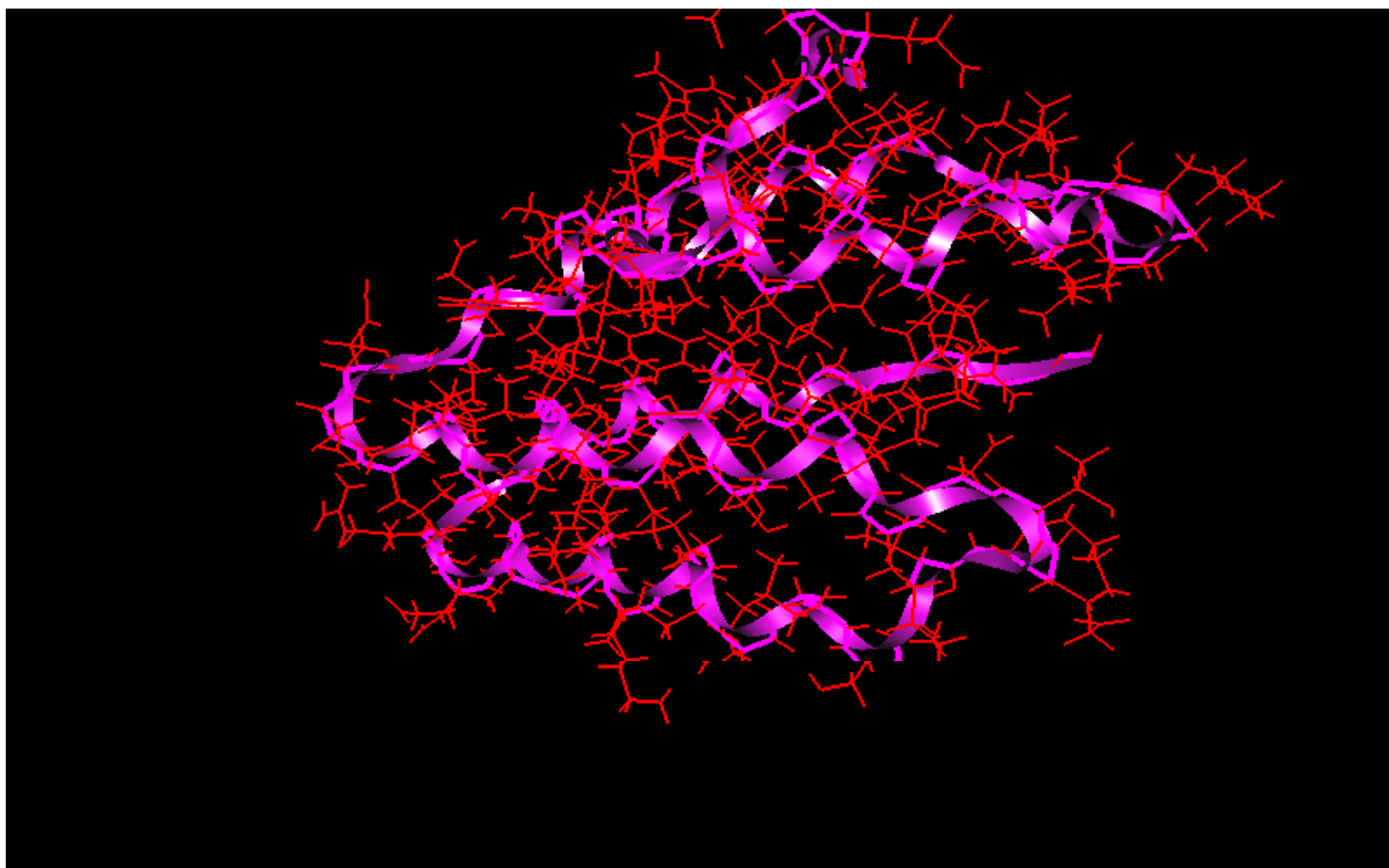
The Cosmic Simulator is the concept of providing an integrated framework in which component simulations can be linked together to provide a coherent, end-to-end, history of the Cosmos.



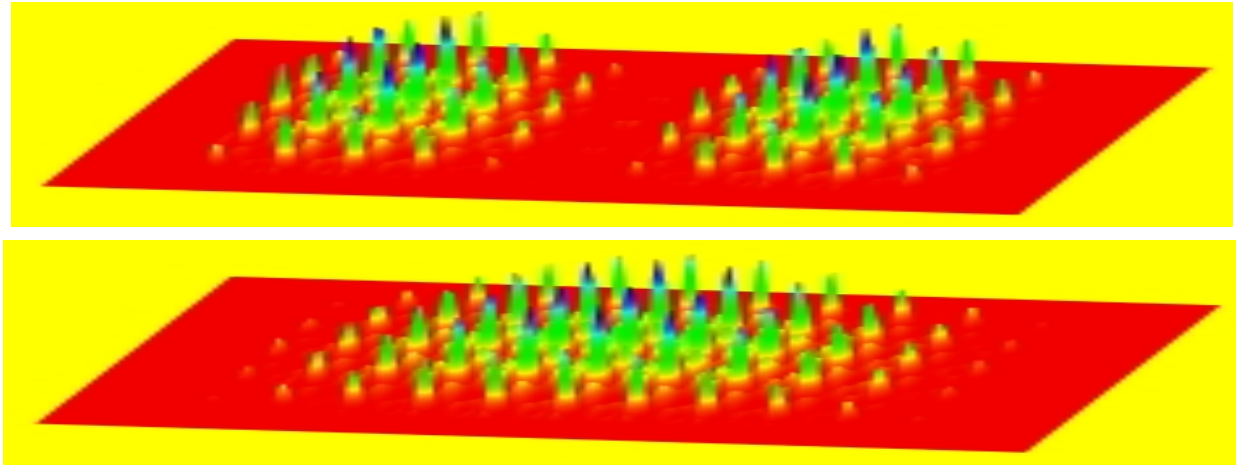
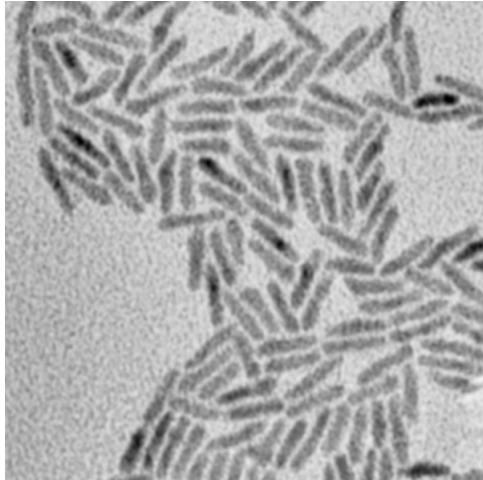
South Pole



Computational Steering of Protein Folding



Polarization of CdSe quantum rods



CdSe quantum rods The electron wavefunctions of a quantum rods

- The electronic structures and optical properties change with the shape of the quantum rods.
- The thousand atom quantum rods can be calculated using the planewave pseudopotential method and the NERSC supercomputers.
- Programs exist at NERSC to calculate such nanosystems and compare with experimental electronic and optical results.

The Future

more resources:

**no limits to growth in demand for
supercomputer resources seen**

better integration:

**computational science and engineering will
become recognized as discipline**

next level simulation science:

**large scale simulation environments will
emerge that allow computer simulation at
unprecedented scale**

Summary



- NERSC is one of the largest open, unclassified supercomputer centers world wide, and is one of the largest “data centers” in science
- CRD has unique capabilities in
 - Distributed computing and grid middleware
 - Data management
 - Cybersecurity
 - Imaging and Visualization
 - Scientific Computing and Numerical Methods
 - Applied Mathematics
 - Applied Computer Science Research
- Both leverage DOE/SC investment through
 - Collaborations in a DOE/SC multipurpose lab
 - International Collaborations
 - University Collaborations

